

ON THE PROPERTIES OF α -UNCHAINING SINGLE LINKAGE HIERARCHICAL CLUSTERING

A. MARTÍNEZ-PÉREZ

ABSTRACT. In the election of a hierarchical clustering method, theoretic properties may give some insight to determine which method is the most suitable to treat a clustering problem. Herein, we study some basic properties of two hierarchical clustering methods: α -unchaining single linkage or $SL(\alpha)$ and a modified version of this one, $SL^*(\alpha)$. We compare the results with the properties satisfied by the classical linkage-based hierarchical clustering methods.

Keywords: Hierarchical clustering, single linkage, chaining effect, weakly unchaining, α -bridge-unchaining.

E-mail: alvaro.martinezperez@uclm.es

Address: Departamento de Análisis Económico y Finanzas. Universidad de Castilla-La Mancha. Avda. Real Fábrica de Seda, s/n. 45600. Talavera de la Reina. Toledo. Spain

CONTENTS

1. Introduction	1
2. Background and notation	3
3. Single linkage hierarchical clustering	6
3.1. Characterization of SL	6
3.2. Stability of SL	8
4. Basic properties of $SL(\alpha)$ and $SL^*(\alpha)$	8
5. Conclusions	13
References	14

1. INTRODUCTION

Kleinberg discussed in [7] the problem of clustering in an axiomatic way. He proposed a few basic properties that any clustering scheme should hold. Let $\mathcal{P}(X)$ denote the set of all possible partitions of X . Fix a clustering method \mathfrak{T} so that $\mathfrak{T}(X) = \Pi \in \mathcal{P}(X)$. The properties proposed by Kleinberg were:

- **Scale invariance:** For all $\alpha > 0$, $\mathfrak{T}(X, \alpha \cdot d) = \Pi$
- **Richness:** Given a finite set X , for every $\Pi \in \mathcal{P}(X)$ there exists a metric d_Π on X such that $\mathfrak{T}(X, d_\Pi) = \Pi$.
- **Consistency:** Let $\Pi = \{B_1, \dots, B_n\}$. Let d' be any metric on X such that
 - 1) for all $x, x' \in B_i$, $d'(x, x') \leq d(x, x')$ and
 - 2) for all $x \in B_i$, $x' \in B_j$, $i \neq j$, $d'(x, x') \geq d(x, x')$.
 Then, $\mathfrak{T}(X, d') = \Pi$.

The author was partially supported by MTM-2009-07030.

Then, he proved that no standard clustering scheme satisfying this conditions simultaneously can exist. This does not mean that defining a clustering function is impossible. The impossibility only holds when the unique input in the algorithm is the space and the set of distances. It can be avoided including, for example, the number of clusters to be obtained as part of the input. See [1] and [11].

Carlsson and Mémoli studied in [4] the analogous problem for hierarchical clustering methods taking as input a finite metric space. They set three basic conditions, see Theorem 3.1, and prove that the unique method satisfying these conditions simultaneously is the well-known single linkage algorithm. The authors prove also that single linkage hierarchical clustering ($SL\ HC$) exhibits some good properties. In particular, it is stable in the Gromov-Hausdorff sense, this is, if two metric spaces are close in the Gromov-Hausdorff metric, then applying the algorithm, the ultrametric spaces obtained are also close in this metric. However, there is a basic weakness in $SL\ HC$ which is the *chaining effect* which can be seen as the tendency of the algorithm to merge two blocks when the minimal distance between them is small ignoring everything else in the distribution.

In [9] we tried to offer some solution to this effect. We proposed a modified version of SL algorithm, α -unchaining single linkage (or $SL(\alpha)$), which shows some sensitivity to the density distribution of the sample and it is capable to distinguish blocks even though the minimal distance between them is small. We also defined a second version of this method, $SL^*(\alpha)$, to detect blocks when they are connected by a chain of points. Then, we studied the unchaining properties of both methods.

Thus, we were able to offer some solution to these chaining effects but, in exchange, we lost some of the good properties of SL . In particular, $SL(\alpha)$ is no longer stable in the Gromov-Hausdorff sense. In fact, as we proved in [10], there is no stable solution to this chaining effect in the range of almost-standard linkage-based HC methods using ℓ^{SL} .

Now, the question is when should we use $SL(\alpha)$ and $SL^*(\alpha)$. Among the large variety of clustering methods the best option usually depends on the particular clustering problem. But how do we choose the most suitable algorithm for the task? Ackerman, Ben David and Loker propose to study significant properties of the clustering functions. See [1] and [2]. The idea is finding abstract significant properties concerning the output of the algorithms which illustrate the difference between applying one clustering method or another. Then, the practitioner should decide which properties are important for the problem under study and choose the algorithm which satisfies them.

In [9] and [10] we proved the chaining, unchaining and stability properties of $SL(\alpha)$ and $SL^*(\alpha)$. Herein, we complete the work by analyzing which of the abstract characteristic properties of SL are also satisfied by these two methods and which properties are lost by adding the unchaining condition.

We start with the characterization of $SL\ HC$ by Carlsson and Mémoli. In the original characterization of SL from [4], properties *I*, *II* and *III* characterize SL , see 3.1. However, we introduce some alternative definitions to offer a better picture of the difference.

We define that a HC method \mathfrak{T} satisfies property *A2* if adding points to the input will never make increase the distance between previous points in the output. This is the case of SL . In fact, we prove that *A2* together with *A1* (the algorithm leaves ultrametric spaces invariant) and *A3* (the distance between two points in the

output is at least the minimal $\varepsilon > 0$ so that there exists a ε -chain between them in the input), offers an alternative characterization of SL . See corollary 3.7.

Properties $A1$ and $A3$ are trivially satisfied by many algorithms, in particular $SL(\alpha)$, $SL^*(\alpha)$, complete linkage (CL) or average linkage (AL). Thus, $A2$ illustrates the difference between SL and other methods as those mentioned above. Also, considering the original characterization from [4], it is trivial to check that $SL(\alpha)$, $SL^*(\alpha)$, AL and CL satisfy I and III but not II and therefore, property II can be used to distinguish those algorithms from SL . However, since II implies $A2$, we believe that $A2$ is a better option for the task.

We also prove that other basic properties as being permutation invariant or rich are satisfied by all of them.

The results obtained in [9], [10] and herein are summarized in Table 1.

2. BACKGROUND AND NOTATION

A dendrogram over a finite set is a nested family of partitions. This is usually represented as a rooted tree.

Let $\mathcal{P}(X)$ denote the collection of all partitions of a finite set $X = \{x_1, \dots, x_n\}$. Then, a dendrogram can also be described as a map $\theta: [0, \infty) \rightarrow \mathcal{P}(X)$ such that:

1. $\theta(0) = \{\{x_1\}, \{x_2\}, \dots, \{x_n\}\}$,
2. there exists T such that $\theta(t) = X$ for every $t \geq T$,
3. if $r \leq s$ then $\theta(r)$ refines $\theta(s)$,
4. for all r there exists $\varepsilon > 0$ such that $\theta(r) = \theta(t)$ for $t \in [r, r + \varepsilon]$.

Notice that conditions 2 and 4 imply that there exist $t_0 < t_1 < \dots < t_m$ such that $\theta(r) = \theta(t_{i-1})$ for every $r \in [t_{i-1}, t_i)$, $i = 0, 1, \dots, m$ and $\theta(r) = \theta(t_m) = \{X\}$ for every $r \in [t_m, \infty)$.

For any partition $\{B_1, \dots, B_k\} \in \mathcal{P}(X)$, the subsets B_i are called *blocks*.

Let $\mathcal{D}(X)$ denote the collection of all possible dendrograms over a finite set X . Given some $\theta \in \mathcal{D}(X)$, let us denote $\theta(t) = \{B_1^t, \dots, B_{k(t)}^t\}$. Therefore, the nested family of partitions is given by the corresponding partitions at t_0, \dots, t_m , this is, $\{B_1^{t_i}, \dots, B_{k(t_i)}^{t_i}\}$, $i = 0, \dots, m$.

An *ultrametric space* is a metric space (X, d) such that $d(x, y) \leq \max\{d(x, z), d(z, y)\}$ for all $x, y, z \in X$. Given a finite metric space X let $\mathcal{U}(X)$ denote the set of all ultrametrics over X .

There is a well known equivalence between trees and ultrametrics. See [6] and [8] for a complete exposition of how to build categorical equivalences between them. In particular, this may be translated into an equivalence between dendrograms and ultrametrics:

Thus, a hierarchical clustering method \mathfrak{T} can be presented as an algorithm whose output is a dendrogram or an ultrametric space. Let $\mathfrak{T}_{\mathcal{D}}(X, d)$ denote the dendrogram obtained by applying \mathfrak{T} to a metric space (X, d) and $\mathfrak{T}_{\mathcal{U}}(X, d)$ denote the corresponding ultrametric space.

In [4] the authors use a recursive procedure to redefine SL HC , average linkage (AL) and complete linkage (CL) hierarchical clustering. The main advantage of this procedure is that it allows to merge more than two clusters at the same time. Therefore, AL and CL HC can be made *permutation invariant*, meaning that the result of the hierarchical clustering does not depend on the order in which the points are introduced in the algorithm. In [9] we gave an alternative presentation of this

recursive procedure as a first step to define $SL(\alpha)$ and $SL^*(\alpha)$. Let us recall here, for completeness, this presentation.

For $x, y \in X$ and any (standard) clustering C of X , $x \sim_C y$ if x and y belong to the same cluster in C and $x \not\sim_C y$, otherwise.

Two (standard) clusterings $C = (C_1, \dots, C_k)$ of (X, d) and $C' = (C'_1, \dots, C'_k)$ of (X', d') are isomorphic clusterings, denoted $(C, d) \cong (C', d')$, if there exists a bijection $\phi : X \rightarrow X'$ such that for all $x, y \in X$, $d(x, y) = d'(\phi(x), \phi(y))$ and $x \sim_C y$ if and only if $\phi(x) \sim_{C'} \phi(y)$.

Definition 2.1. A linkage function is a function

$$\ell : \{(X_1, X_2, d) \mid d \text{ is a distance function over } X_1 \cup X_2\} \rightarrow \mathbb{R}^+$$

such that,

1. ℓ is representation independent: For all (X_1, X_2) and (X'_1, X'_2) , if $(X_1, X_2, d) \cong (X'_1, X'_2, d')$ (i.e., they are clustering-isomorphic), then $\ell(X_1, X_2, d) = \ell(X'_1, X'_2, d')$.
2. ℓ is monotonic: For all (X_1, X_2) if d' is a distance function over $X_1 \cup X_2$ such that for all $x \sim_{\{X_1, X_2\}} y$, $d(x, y) = d'(x, y)$ and for all $x \not\sim_{\{X_1, X_2\}} y$, $d(x, y) \leq d'(x, y)$ then $\ell(X_1, X_2, d') \geq \ell(X_1, X_2, d)$.
3. Any pair of clusters can be made arbitrarily distant: For any pair of data sets (X_1, d_1) , (X_2, d_2) , and any r in the range of ℓ , there exists a distance function d that extends d_1 and d_2 such that $\ell(X_1, X_2, d) > r$.

For technical reasons, it is usually assumed that a linkage function has a countable range. Say, the set of nonnegative algebraic real numbers.

Some standard choices for ℓ are:

- Single linkage: $\ell^{SL}(B, B') = \min_{(x, x') \in B \times B'} d(x, x')$
- Complete linkage: $\ell^{CL}(B, B') = \max_{(x, x') \in B \times B'} d(x, x')$
- Average linkage: $\ell^{AL}(B, B') = \frac{\sum_{(x, x') \in B \times B'} d(x, x')}{\#(B) \cdot \#(B')}$ where $\#(X)$ denotes the cardinality of the set X .

Let (X, d) be a finite metric space where $X = \{x_1, \dots, x_n\}$. Let L denote a family of linkage functions on X and fix some linkage function $\ell \in L$. Then, let $\mathfrak{T}_{\mathcal{D}}(X, d) = \theta^\ell$ be as follows:

1. Let $\Theta_0 := \{x_1, \dots, x_n\}$ and $R_0 = 0$.
2. For every $i \geq 1$, while $\Theta_{i-1} \neq \{X\}$, let $R_i := \min\{\ell(B, B') \mid B, B' \in \Theta_{i-1}, B \neq B'\}$. Then, let $G_{R_i}^\ell$ be a graph whose vertices are the blocks of Θ_{i-1} and such that there is an edge joining B and B' if and only if $\ell(B, B') \leq R_i$.
3. Consider the equivalence relation $B \sim_{\ell, R} B'$ if and only if B, B' are in the same connected component of $G_{R_i}^\ell$. Then, $\Theta_i = \frac{\Theta_{i-1}}{\sim_{\ell, R_i}}$.
4. Finally, let $\theta^\ell : [0, \infty) \rightarrow \mathcal{P}(X)$ be such that $\theta^\ell(r) := \Theta_{i(r)}$ with $i(r) := \max\{i \mid R_i \leq r\}$.

In [10] the methods defined by applying this algorithm for some linkage function ℓ are called *standard linkage-based HC* methods.

Let us now recall the definition of $SL(\alpha)$ and $SL^*(\alpha)$. Further explanations, figures and easy examples of applications of these methods can be found in [9].

Given a finite metric space (X, d) , let $F_t(X, d)$ be the Rips (or Vietoris-Rips) complex of (X, d) . Let us recall that the Rips complex of a metric space (X, d) is a simplicial complex whose vertices are the points of X and $[v_0, \dots, v_k]$ is a simplex

of $F_t(X, d)$ if and only if $d(v_i, v_j) \leq t$ for every i, j . Given any subset $Y \subset X$, by $F_t(Y)$ we refer to the subcomplex of $F_t(X)$ defined by the vertices in Y . A simplex $[v_0, \dots, v_k]$ has dimension k . The dimension of a simplicial complex is the maximal dimension of its simplices.

Let $X = \{x_1, \dots, x_n\}$. Let $d_{ij} := d(x_i, x_j)$ and $D := \{t_i : 0 \leq i \leq m\} = \{d_{ij} : 1 \leq i, j \leq n\}$ with $t_i < t_j \forall i < j$ where " $<$ " denotes the order of the real numbers. Clearly, $t_0 = 0$.

Let the dendrogram defined by $SL(\alpha)$, $\mathfrak{T}_D^{SL(\alpha)}(X, d) = \theta_{X, \alpha}$ or simply θ_α , be as follows:

- 1) Let $\theta_\alpha(0) := \{\{x_1\}, \dots, \{x_n\}\}$ and $\theta_\alpha(t) := \theta_\alpha(0) \forall t < t_1$. Now, for every i , given $\theta_\alpha[t_{i-1}, t_i) = \theta_\alpha(t_{i-1}) = \{B_1, \dots, B_m\}$, we define recursively θ_α on the interval $[t_i, t_{i+1})$ as follows:
 - 2) Let $G_\alpha^{t_i}$ be a graph with vertices $\mathcal{V}(G_\alpha^{t_i}) := \{B_1, \dots, B_m\}$ and edges $\mathcal{E}(G_\alpha^{t_i}) := \{B_j, B_k\}$ such that the following conditions hold:
 - i) $\min\{d(x, y) \mid x \in B_j, y \in B_k\} \leq t_i$.
 - ii) there is a simplex $\Delta \in F_{t_i}(B_j \cup B_k)$ such that $\Delta \cap B_j \neq \emptyset$, $\Delta \cap B_k \neq \emptyset$ and $\alpha \cdot \dim(\Delta) \geq \min\{\dim(F_{t_i}(B_j)), \dim(F_{t_i}(B_k))\}$.
 - 3) Let us define a relation, $\sim_{t_i, \alpha}$ as follows.

Let $B_j \sim_{t_i, \alpha} B_k$ if B_j, B_k belong to the same connected component of the graph $G_\alpha^{t_i}$. Then, $\sim_{t_i, \alpha}$ induces an equivalence relation.
 - 4) For every $t \in [t_i, t_{i+1})$, $\theta_\alpha(t) := \theta_\alpha(t_{i-1}) / \sim_{t_i, \alpha}$.

This construction is generalized in [10] to define the class of *almost-standard linkage-based HC* methods.

Remark 2.2. Notice that if two points x, x' belong to the same block of $\theta_\alpha(t_i)$ then, necessarily, there exists a t_i -chain, $x = x_0, x_1, \dots, x_n = x'$ joining them. In particular, if $x_j \in B_j \in \theta_\alpha(t_{i-1})$, $j = 0, \dots, n$, the corresponding edges $\{B_{j-1}, B_j\}$, $1 \leq j \leq n$, satisfy condition ii. This is immediate by construction.

Let the dendrogram defined by $SL^*(\alpha)$, $\mathfrak{T}_D^{SL^*(\alpha)}(X, d) = \theta_{X, \alpha}^*$ or simply θ_α^* , be as follows:

- 1) Let $\theta_\alpha^*(0) := \{\{x_1\}, \dots, \{x_n\}\}$ and $\theta_\alpha^*(t) := \theta_\alpha^*(0) \forall t < t_1$.

Now, given $\theta_\alpha^*[t_{i-1}, t_i) = \theta_\alpha^*(t_{i-1}) = \{B_1, \dots, B_m\}$, we define recursively θ_α^* on the interval $[t_i, t_{i+1})$ as follows:
- 2) Let $G_\alpha^{t_i}$ be a graph with vertices $\mathcal{V}(G_\alpha^{t_i}) := \{B_1, \dots, B_m\}$ and edges $\mathcal{E}(G_\alpha^{t_i}) := \{B_j, B_k\}$ such that the following conditions hold:
 - i) $\min\{d(x, y) \mid x \in B_j, y \in B_k\} \leq t_i$.
 - ii) there is a simplex $\Delta \in F_{t_i}(B_j \cup B_k)$ such that $\Delta \cap B_j \neq \emptyset$, $\Delta \cap B_k \neq \emptyset$ and $\alpha \cdot \dim(\Delta) \geq \min\{\dim(F_{t_i}(B_j)), \dim(F_{t_i}(B_k))\}$.

By an abuse of the notation, we may write B to refer both to the block of $\theta(t_{i-1})$ and to the vertex of $G_\alpha^{t_i}$.
- 3) Let us define a relation, $\sim_{t_i, \alpha}$ between the blocks as follows.

Let $cc(G_\alpha^{t_i})$ be the set of connected components of the graph $G_\alpha^{t_i}$. Let $A \in cc(G_\alpha^{t_i})$ with $A = \{B_{j_1}, \dots, B_{j_r}\}$.
Let us call *big blocks* of A those blocks such that

$$(1) \quad \alpha \cdot \#(B_{j_k}) \geq \max_{1 \leq l \leq r} \{\#(B_{j_l})\}.$$

The rest of blocks of A are called *small blocks*.

Let $H_\alpha(A)$ be the subgraph of A whose vertices are the big blocks and $S_\alpha(A)$ be the subgraph of A whose vertices are the small blocks.

Then, $B_{j_k} \sim_{t_i, \alpha} B_{j_{k'}}$ if one of the following conditions holds:

- iii) $\exists C \in cc(H_\alpha(A))$ such that $B_{j_k}, B_{j_{k'}} \in C$.
- iv) $B_{j_k} \in C \in cc(H_\alpha(A))$, $B_{j_{k'}} \in C' \in cc(S_\alpha(A))$ and there is no big block in $A \setminus C$ adjacent to any block in C' .

Then, $\sim_{t_i, \alpha}$ induces an equivalence relation whose classes are contained in the connected components of $G_\alpha^{t_i}$.

- 4) For every $t \in [t_i, t_{i+1})$, $\theta_\alpha^*(t) := \theta_\alpha^*(t_{i-1}) / \sim_{t_i, \alpha}$.

Remark 2.3. At step iii, if $H_\alpha(A)$ is connected, then $B_{j_1} \cup \dots \cup B_{j_r}$ defines a block of $\theta_\alpha(t_i)$.

Remark 2.4. Notice that Remark 2.2 still applies. In fact, if two points x, x' belong to the same block of $\theta_\alpha^*(t_i)$ then, necessarily, there exists a t_i -chain, $x = x_0, x_1, \dots, x_n = x'$ joining them so that if $x_j \in B_j \in \theta_\alpha^*(t_{i-1})$, $j = 0, \dots, n$, the corresponding edges $\{B_{j-1}, B_j\}$, $1 \leq j \leq n$, satisfy condition ii.

3. SINGLE LINKAGE HIERARCHICAL CLUSTERING

In this section we recall some basic properties and the characterization of SL HC from [4]. We also propose some alternatives. Our first intention is to find significant properties to compare SL and $SL(\alpha)$.

3.1. Characterization of SL . Carlsson and Mémoli provided the following axiomatic characterization of SL HC :

Let us recall that given a finite metric space (X, d) , $sep(X, d) := \min_{x \neq x'} d(x, x')$.

Theorem 3.1. [4, Theorem 18] Let \mathfrak{T} be a hierarchical clustering method such that:

- (I) $\mathfrak{T}_U\left(\{p, q\}, \begin{pmatrix} \delta & 0 \\ 0 & \delta \end{pmatrix}\right) = \left(\{p, q\}, \begin{pmatrix} \delta & 0 \\ 0 & \delta \end{pmatrix}\right)$ for all $\delta > 0$.
- (II) Given two finite metric spaces X, Y and $\phi: X \rightarrow Y$ such that $d_X(x, x') \geq d_Y(\phi(x), \phi(x'))$ for all $x, x' \in X$, then

$$u_X(x, x') \geq u_Y(\phi(x), \phi(x'))$$

also holds for all $x, x' \in X$, where $\mathfrak{T}_U(X, d_X) = (X, u_X)$ and $\mathfrak{T}_U(Y, d_Y) = (Y, u_Y)$.

- (III) For any metric space (X, d) ,

$$u(x, x') \geq sep(X, d) \text{ for all } x \neq x' \in X$$

where $\mathfrak{T}_U(X, d) = (X, u)$.

Then, \mathfrak{T} is exactly single linkage hierarchical clustering.

Notation: For the particular case of SL HC , if there is no need to distinguish the metric space, let us denote $\mathfrak{T}_D^{SL}(X, d) = \theta_{SL}$ and $\mathfrak{T}_U^{SL}(X, d) = (X, u_{SL})$.

Notation: Given two metrics d, d' defined on a set X , let us denote $d \leq d'$ if $d(x, x') \leq d'(x, x') \forall x, x' \in X$.

The following propositions follow immediately from the proof of [4, Theorem 18].

Proposition 3.2. For any metric space (X, d) , if \mathfrak{T} satisfies conditions II and III, then $u \geq u_{SL}$.

This is, the ultrametric distance between two points is at least the minimal length ε for which there is a ε -chain joining them.

It is readily seen that if $u \geq u_{SL}$, then \mathfrak{T} satisfies *III*.

Proposition 3.3. *If \mathfrak{T} satisfies conditions I and II, then $u_{SL} \geq u$.*

In fact, Proposition 3.3 can be improved introducing the following condition.

A2) Let (Y, d) be a metric space and $X \subset Y$. If $i: X \rightarrow Y$ is the inclusion map, then $u_X(x, x') \geq u_Y(i(x), i(x'))$.

This is, by adding points to the space we may make the ultrametric distance smaller but never bigger. Clearly, $II \Rightarrow A2$. The proof of Proposition 3.3, [4], can be trivially adapted to obtain the following.

Proposition 3.4. *If \mathfrak{T} satisfies conditions I and A2, then $u_{SL} \geq u$.*

Proof. Let $x, x' \in (X, d)$ such that $u_{SL}(x, x') = \delta$. Then, there exists a δ -chain $x = x_0, x_1, \dots, x_n = x'$ such that $\max_i d(x_{i-1}, x_i) = \delta$. By I, if $X_i = \{x_{i-1}, x_i\}$, $\mathfrak{T}(X_i, d|_{X_i}) = (X_i, d|_{X_i})$ and $u_{X_i}(x_{i-1}, x_i) \leq \delta$. Then, by A2, $u(x_{i-1}, x_i) \leq u_{X_i}(x_{i-1}, x_i) \leq \delta$ and, by the properties of the ultrametric, $u(x, x') \leq \delta$. \square

Another natural condition to ask on a hierarchical clustering method is leaving invariant any ultrametric space:

A1) If (X, d) is an ultrametric space, then $u(x, y) = d(x, y)$.

This is, applying the hierarchical clustering method to an ultrametric space we obtain the same ultrametric space.

Also, it can be readily seen that $SL\ HC$ satisfies A1:

Proposition 3.5. *If (X, d) is an ultrametric space, then $u_{SL}(x, y) = d(x, y)$ for every $x, y \in X$.*

Proof. By definition, it is clear that $u_{SL}(x, y) \leq d(x, y)$ for every $x, y \in X$.

Let us see that, if (X, d) is an ultrametric space, then $u_{SL}(x, y) \geq d(x, y)$. $u_{SL}(x, y) = \inf\{t \mid \text{there exists a } t\text{-chain joining } x \text{ to } y\}$. Suppose $u_{SL}(x, y) = t$ and let $x = x_0, x_1, \dots, x_n = y$ be a t -chain joining x to y . By the properties of the ultrametric, $d(x_{i-1}, x_{i+1}) \leq \max\{d(x_{i-1}, x_i), d(x_i, x_{i+1})\} \leq t$ for every $1 \leq i \leq n-1$. Therefore, $d(x, y) \leq t$ and $u_{SL}(x, y) \geq d(x, y)$. \square

Richness property for HC methods can be defined in the same way Kleinberg did for standard clustering. Thus, a HC method \mathfrak{T} satisfies *richness property* if given a finite set X , for every $\theta \in \mathcal{D}(X)$ there exists a metric d_θ on X such that $\mathfrak{T}_D(X, d_\theta) = \theta$.

Corollary 3.6. *\mathfrak{T}^{SL} satisfies richness property.*

It is trivial to check that $A1 \Rightarrow I$ (and $A3 \Rightarrow III$). Therefore, by Proposition 3.4, we obtain also the following alternative characterization of $SL\ HC$.

Corollary 3.7. *Let \mathfrak{T} be a hierarchical clustering method such that:*

- A1) *If (X, d) is an ultrametric space, then $u_X(x, y) = d(x, y)$.*
- A2) *Let (Y, d) be a metric space and $X \subset Y$. If $i: X \rightarrow Y$ is the inclusion map, then $u_X(x, x') \geq u_Y(i(x), i(x'))$.*
- A3) *$u \geq u_{SL}$.*

Then, \mathfrak{T} is exactly $SL\ HC$.

3.2. Stability of SL . Let us recall the definition of Gromov-Hausdorff distance from [3]. See also [5].

Let (X, d_X) and (Y, d_Y) two metric spaces. A correspondence (between A and B) is a subset $R \in A \times B$ such that

- $\forall a \in A$, there exists $b \in B$ s.t. $(a, b) \in R$
- $\forall b \in B$, there exists $a \in A$ s.t. $(a, b) \in R$

Let $\mathcal{R}(A, B)$ denote the set of all possible correspondences between A and B .

Let $\Gamma_{X,Y}: X \times Y \times X \times Y \rightarrow \mathbb{R}^+$ given by

$$(x, y, x', y') \mapsto |d_X(x, x') - d_Y(y, y')|.$$

Then, the *Gromov-Hausdorff distance* between X and Y is:

$$d_{\mathcal{GH}}(X, Y) := \frac{1}{2} \inf_{R \in \mathcal{R}(X, Y)} \sup_{(x, y), (x', y') \in R} \Gamma_{X, Y}(x, y, x', y').$$

The *Gromov-Hausdorff metric* gives a notion of distance between metric spaces. One of the advantages of this metric is that it is well defined for metric spaces of different cardinality. In [4] this metric is used to prove that \mathfrak{T}^{SL} holds some stability under small perturbations on the metric. The authors prove that if two metric spaces are close (in the Gromov-Hausdorff metric) then the corresponding ultrametric spaces obtained as output of the algorithm are also close. In [10] we studied Gromov-Hausdorff stability of linkage-based *HC* methods defining the following conditions.

Notation: Let (\mathcal{M}, d_{GH}) denote the set of finite metric spaces with the Gromov-Hausdorff metric and (\mathcal{U}, d_{GH}) denote the set of finite ultrametric spaces with the Gromov-Hausdorff metric.

Definition 3.8. A *HC method* \mathfrak{T} is semi-stable in the Gromov-Hausdorff sense if for any sequence of finite metric spaces $((X_k, d_k))_{k \in \mathbb{N}}$ in (\mathcal{M}, d_{GH}) such that $\lim_{k \rightarrow \infty} (X_k, d_k) = (U, d) \in \mathcal{U}$ then $\lim_{k \rightarrow \infty} \mathfrak{T}_{\mathcal{U}}(X_k, d_k) = \mathfrak{T}_{\mathcal{U}}(U, d)$.

Definition 3.9. A *HC method* \mathfrak{T} is stable in the Gromov-Hausdorff sense if

$$\mathfrak{T}_{\mathcal{U}}: (\mathcal{M}, d_{GH}) \rightarrow (\mathcal{U}, d_{GH})$$

is continuous.

A hierarchical clustering method is said to be *permutation invariant* if it yields the same dendrogram under permutation of the points in the sample this is, if the output of the algorithm does not depend on the order by which the data is introduced. Although this is not the easiest way to check this property, it may be noticed that being stable in the Gromov-Hausdorff sense implies being permutation invariant.

The following result is a consequence of [4, Proposition 26].

Proposition 3.10. *SL HC is stable in the Gromov-Hausdorff sense. In particular, it is semi-stable and permutation invariant.*

4. BASIC PROPERTIES OF $SL(\alpha)$ AND $SL^*(\alpha)$

In this section, we study some basic properties on $SL(\alpha)$ and $SL^*(\alpha)$. In particular, we check those seen at Section 3.

The following result is clear from the definition.

Proposition 4.1. $SL(\alpha)$ and $SL^*(\alpha)$ are permutation invariant algorithms.

Proposition 4.2. Let (X, d) be a finite metric space with $X = \{x_1, \dots, x_n\}$. If $\alpha \geq \frac{n-2}{2}$, then $\mathfrak{T}^{SL}(X) = \mathfrak{T}^{SL(\alpha)}(X)$.

Proof. Let $\mathfrak{T}_D^{SL}(X) = \theta_{SL}$, $\mathfrak{T}_D^{SL(\alpha)}(X) = \theta_\alpha$.

We know that $\theta_{SL}(t_0) = \theta_\alpha(t_0)$. Suppose $\theta_{SL}(t_{i-1}) = \theta_\alpha(t_{i-1})$.

Let us see that for $\alpha \geq \frac{n-2}{2}$, condition i already implies ii and the edges of the graph $G_\alpha^{t_i}$ are those defined by condition i . Let B_1, B_2 two blocks in $\theta_\alpha(t_{i-1})$ such that $\min\{d(x, y) \mid x \in B_1, y \in B_2\} \leq t_i$. For any simplex Δ , $\alpha \cdot \dim(\Delta) \geq \alpha$ and $\min\{\#(B_1), \#(B_2)\} \leq \frac{n}{2}$. Since $\alpha \geq \frac{n-2}{2}$, $\alpha \cdot \dim(\Delta) \geq \alpha \geq \min\{\#(B_1) - 1, \#(B_2) - 1\} \geq \min\{\dim(F_{t_i}(B_1)), \dim(F_{t_i}(B_2))\}$.

Then, $\theta_{SL}(t_i) = \theta_\alpha(t_i)$. \square

Proposition 4.3. Let (X, d) be a finite metric space with $X = \{x_1, \dots, x_n\}$. If $\alpha \geq n - 1$, then $\mathfrak{T}^{SL}(X) = \mathfrak{T}^{SL^*(\alpha)}(X)$.

Proof. Let $\mathfrak{T}_D^{SL}(X) = \theta_{SL}$ and $\mathfrak{T}_D^{SL^*(\alpha)}(X) = \theta_\alpha^*$.

We know that $\theta_{SL}(t_0) = \theta_\alpha^*(t_0)$. Suppose $\theta_{SL}(t_{i-1}) = \theta_\alpha^*(t_{i-1})$.

As we saw in the proof of Proposition 4.2, since $\alpha \geq n - 1 > \frac{n-2}{2}$, condition i already implies ii and the edges of the graph $G_\alpha^{t_i}$ are those defined by condition i .

Now, let $A = \{B_1, \dots, B_r\}$ be any connected component of $G_\alpha^{t_i}$.

If the subgraph $H_\alpha(A)$ is not connected, then there are at least three blocks $B_{i_1}, B_{i_2}, B_{i_3}$ in A , such that $1 \leq \#(B_{i_1}) < \frac{1}{\alpha} \max_{1 \leq l \leq r} \{\#(B_l)\}$ and $\#(B_{i_2}), \#(B_{i_3}) \geq \frac{1}{\alpha} \max_{1 \leq l \leq r} \{\#(B_l)\}$. Trivially, $\max_{1 \leq l \leq r} \{\#(B_l)\} \leq n - 2$. Hence, there is a contradiction since $1 \leq \frac{n-2}{\alpha} \leq \frac{n-2}{n-1} < 1$.

Thus, $H_\alpha(A)$ is connected and, as we saw in Remark 2.3, all the blocks in A are identified. Therefore, $\theta_\alpha^*(t_i) = \theta_{SL}(t_i)$. \square

Notation: Let X be a finite metric space. Let us recall that if there is no ambiguity on the metric space we denote $\mathfrak{T}_D^{SL}(X) = \theta_{SL}$, $\eta(\theta_{SL}) = u_{SL}$ and $\mathfrak{T}_D^{SL(\alpha)}(X) = \theta_\alpha$. Let us denote $\eta(\theta_\alpha) = u_\alpha$. Similarly, let $\mathfrak{T}_D^{SL^*(\alpha)}(X) = \theta_\alpha^*$ and $\eta(\theta_\alpha^*) = u_\alpha^*$.

Proposition 4.4. $u_{SL} \leq u_\alpha$ and $u_{SL} \leq u_\alpha^*$ for every $\alpha \in \mathbb{N}$ (i.e. $\mathfrak{T}^{SL(\alpha)}$ and $\mathfrak{T}^{SL^*(\alpha)}$ satisfy A3).

Proof. As we saw at remarks 2.2 and 2.4, if two points $x, x' \in X$ belong to the same block of $\theta_\alpha(t)$ (resp. $\theta_\alpha^*(t)$), they belong, in particular, to the same t -component of X and, therefore, to the same block of $\theta_{SL}(t)$. Thus, $u_{SL}(x, x') \leq u_\alpha(x, x')$ (resp. $u_{SL}(x, x') \leq u_\alpha^*(x, x')$). \square

Proposition 4.5. If (X, d) is an ultrametric space, then $\theta_\alpha = \theta_{SL} = \theta_\alpha^*$ for every α .

Proof. By definition, $\theta_\alpha(t_0) = \theta_{SL}(t_0) = \theta_\alpha^*(t_0)$. Suppose $\theta_\alpha(t_{i-1}) = \theta_{SL}(t_{i-1}) = \theta_\alpha^*(t_{i-1}) = \{B_1, \dots, B_n\}$. Let us see that $\theta_\alpha(t_i) = \theta_{SL}(t_i) = \theta_\alpha^*(t_i)$.

Let B_i, B_j be such that $\min\{d(x, y) \mid x \in B_i, y \in B_j\} \leq t_i$. Since B_i, B_j are (t_{i-1}) -components, by the properties of the ultrametric, $d(x, y) \leq t_i$ for every $(x, y) \in B_1 \times B_2$.

Therefore, every pair of points in $B_1 \cup B_2$ define a simplex in $F_{t_i}(B_1 \cup B_2)$ and condition ii holds for every α . Thus, there is an edge defined between B_i and B_j . This proves that $\theta_\alpha = \theta_{SL}$.

Now, let B_i, B_j be two blocks in the same connected component of $G_\alpha^{t_i}$. Then, by the properties of the ultrametric, $\{B_i, B_j\}$ is an edge of $G_\alpha^{t_i}$. Hence, $H_\alpha(G_\alpha^{t_i})$ is connected and, as we saw in Remark 2.3, $\theta_\alpha^*(t_i)$ is defined by the connected components of $G_\alpha^{t_i}$. This proves that $\theta_\alpha^* = \theta_{SL}$. \square

Corollary 4.6. *If (X, d) is an ultrametric space, then $u_\alpha(x, y) = u_\alpha^*(x, y) = d(x, y)$ for every $x, y \in X$.*

Corollary 4.7. *$SL(\alpha)$ and $SL^*(\alpha)$ satisfy A1 and A3 but not A2.*

Notice that if A2 were also satisfied then, by Corollary 3.7, the method would be exactly SL . For an example of how these methods fail to satisfy A2 consider the following example from [9].

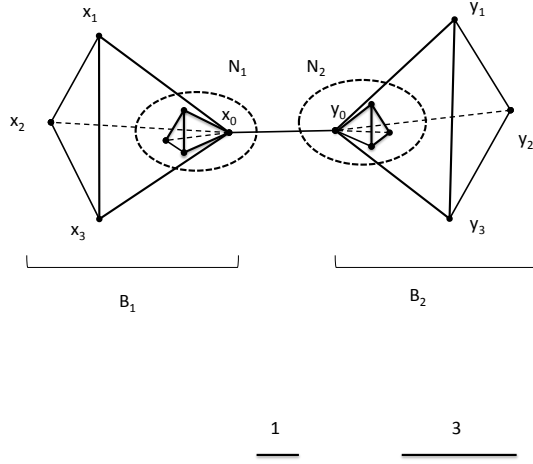


FIGURE 1. $SL(\alpha)$ with $\alpha < 3$ does not satisfy A2.

Example 4.8. *Let (X, d) be the graph from Figure 1.*

Suppose the edges in N_1, N_2 have length 1 and the rest have length 3. The distances between vertices are measured as the minimal length of a path joining them.

Let $Z := \{x_0, y_0\}$ and $d'(x_0, y_0) = 3$. Let $i: (Z, d') \rightarrow (X, d)$ be the inclusion map. It is immediate to check that applying either $SL(\alpha)$ or $SL^(\alpha)$ with $\alpha < 3$ we obtain ultrametric spaces (Z, u_Z) , (X, u_X) such that $u_Z(x_0, y_0) < u_X(x_0, y_0)$.*

Corollary 4.9. *$SL(\alpha)$ and $SL^*(\alpha)$ satisfy richness property.*

As we saw in [10], $SL(\alpha)$ is semi-stable in the Gromov-Hausdorff sense. Unfortunately, most of the good stability properties of SL do not hold. $SL(\alpha)$ and $SL^*(\alpha)$ are not stable in the Gromov-Hausdorff sense (see [10]) and it is not difficult to check that $SL^*(\alpha)$ is not semi-stable in the Gromov-Hausdorff sense. Small perturbations on the distances may affect the dimension of the Rips complex and

to whether or not condition *ii* applies. Also, they may affect the size of the components and yield very different graphs $G_\alpha^{t_i}$. Furthermore, changing the parameter α we may obtain a very different dendrogram. However, all the instability is produced by the unchaining conditions *ii*, *iii* and *iv*. Thus, θ_α and θ_α^* may be compared with θ to, at least, keep track of the undesired effects on the stability introduced with the unchaining conditions.

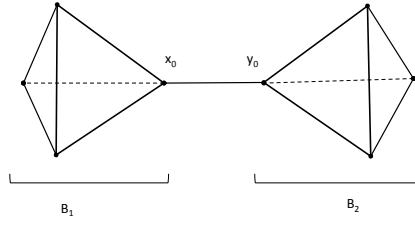


FIGURE 2. A small perturbation in the distance between x_0 and y_0 produces very different dendrograms.

Example 4.10. Let (X, d) be the graph from Figure 2 where every edge has length 1 and let (X, d') be the same graph where $d(x_0, y_0) = 1 + \varepsilon$ for some $\varepsilon > 0$ and the rest of the edges have length 1. Let $\theta_1 = \mathfrak{T}_D^{SL(1)}(X, d)$ and $\theta'_1 = \mathfrak{T}_D^{SL(1)}(X, d')$.

As we saw above, $\theta_1(t) = \{\{x_0\}, \dots, \{x_3\}, \{y_0\}, \dots, \{y_3\}\}$ if $t < 1$ and $\theta_1(1) = \{X\}$. Thus, if $\eta(\theta_1) = u$ it follows that $u(x, y) = 1 \forall x, y \in X$.

If we apply $SL(1)$ to (X, d') we obtain that $\theta'_1(t) = \{\{x_0\}, \dots, \{x_3\}, \{y_0\}, \dots, \{y_3\}\}$ if $t < 1$ and $\theta'_1(t) = \{B_1, B_2\}$ for $1 \leq t < 1 + \varepsilon$. For $1 + \varepsilon \leq t \leq 2$, by condition *ii*, there is no edge in G_1^t between B_1 and B_2 . Thus, $\theta'_1(t) = \{B_1, B_2\}$ for $1 + \varepsilon \leq t < 2 + \varepsilon$. For $t \geq 2 + \varepsilon$, $\theta'_1(t) = X$. Thus, if $\eta(\theta'_1) = u'$ it follows that $u'(x_i, x_j) = 1 \forall x_i, x_j \in B_1$, $u'(y_i, y_j) = 1 \forall y_i, y_j \in B_2$ and $u'(x_i, y_j) = 2 + \varepsilon \forall (x_i, y_j) \in B_1 \times B_2$.

In this case, $d_{GH}((X, d), (X, d')) = \frac{\varepsilon}{2}$ and $d_{GH}((X, u), (X, u')) = \frac{1+\varepsilon}{2}$. Therefore, $SL(\alpha)$ is not stable in the Gromov-Hausdorff sense.

Also, it is unstable under the change of the parameter α .

Example 4.11. Let (X, d') be the graph with the metric defined in Example 4.10.

As we just saw, if $\eta(\theta'_1) = u'$ it follows that $u'(x_i, x_j) = 1 \forall x_i, x_j \in B_1$, $u'(y_i, y_j) = 1 \forall y_i, y_j \in B_2$ and $u'(x_i, y_j) = 2 + \varepsilon \forall (x_i, y_j) \in B_1 \times B_2$.

If we apply $SL(3)$ to (X, d') we obtain that $\theta'_3(t) = \{\{x_0\}, \dots, \{x_3\}, \{y_0\}, \dots, \{y_3\}\}$ if $t < 1$ and $\theta'_3(t) = \{B_1, B_2\}$ for $1 \leq t < 1 + \varepsilon$. For $1 + \varepsilon \leq t \leq 2$, since $\alpha = 3$ there is an edge in G_3^t between B_1 and B_2 . Thus, $\theta'_3(t) = \{X\}$ for $1 + \varepsilon \leq t$. Hence, if $\eta(\theta'_3) = u''$ it follows that $u''(x_i, x_j) = 1 \forall x_i, x_j \in B_1$, $u''(y_i, y_j) = 1 \forall y_i, y_j \in B_2$ and $u''(x_i, y_j) = 1 + \varepsilon \forall (x_i, y_j) \in B_1 \times B_2$.

Therefore, $d_{GH}((X, u'), (X, u'')) = \frac{1}{2}$.

One may wonder if given $\alpha > \alpha'$ anything can be told about the corresponding dendrograms. In particular, given $\mathfrak{T}_{\mathcal{U}}^{SL^*(\alpha)}(X, d) = u_{\alpha}^*$ and $\mathfrak{T}_{\mathcal{U}}^{SL^*(\alpha')}(X, d) = u_{\alpha'}^*$, it is natural to ask if $u_{\alpha}^* \leq u_{\alpha'}^*$ or $u_{\alpha'}^* \leq u_{\alpha}^*$. This need not be true. In fact, it may fail by conditions *iii* and *iv*, see Example 4.12, or by condition *ii*, see Example 4.13.

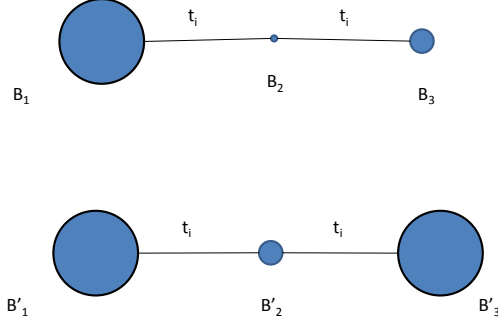


FIGURE 3. A change on the parameter α may have different effects on the chaining through smaller blocks.

Example 4.12. Let $\alpha > \alpha'$. Suppose that $\theta_{\alpha}^*(t_{i-1}) = \theta_{\alpha'}^*(t_{i-1}) = \{B_1, B_2, B_3\}$. See the example above from Figure 3. Now, suppose that conditions *i*, *ii* define edges $\{B_1, B_2\}$ and $\{B_2, B_3\}$ but not $\{B_1, B_3\}$ in both $G_{\alpha}^{t_i}$ and $G_{\alpha'}^{t_i}$.

Suppose that $\max_{1 \leq l \leq 3} \{\#(B_l)\} = \#(B_1)$. Also, let us suppose that $\alpha \cdot \#(B_2) < \#(B_1)$, $\alpha' \cdot \#(B_2) < \#(B_1)$, $\alpha \cdot \#(B_3) \geq \#(B_1)$ but $\alpha' \cdot \#(B_3) < \#(B_1)$. In this case, there is a unique connected component $A = \{B_1, B_2, B_3\}$ and $H_{\alpha'}(A) = \{B_1\}$ is connected while $H_{\alpha}(A) = \{B_1, B_3\}$ is not connected. Thus, $\theta_{\alpha}^*(t_i) = \{B_1, B_2, B_3\}$ and $\theta_{\alpha'}^*(t_i) = \{B_1 \cup B_2 \cup B_3\} = \{X\}$.

Suppose that $\theta_{\alpha}^*(t_{i-1}) = \theta_{\alpha'}^*(t_{i-1}) = \{B'_1, B'_2, B'_3\}$. See the example below from Figure 3. Now, suppose that conditions *i*, *ii* define edges $\{B'_1, B'_2\}$ and $\{B'_2, B'_3\}$ but not $\{B'_1, B'_3\}$ in both $G_{\alpha}^{t_i}$ and $G_{\alpha'}^{t_i}$.

Suppose that $\max_{1 \leq l \leq 3} \{\#(B'_l)\} = \#(B'_1)$. Let us suppose that $\alpha \cdot \#(B'_3) > \#(B'_1)$, $\alpha' \cdot \#(B'_3) > \#(B'_1)$, $\alpha \cdot \#(B'_2) \geq \#(B'_1)$ but $\alpha' \cdot \#(B'_2) < \#(B'_1)$. In this case, there is a unique connected component $A = \{B'_1, B'_2, B'_3\}$, $H_{\alpha}(A)$ has vertices B'_1, B'_2, B'_3 and it is connected while $H_{\alpha'}(A) = \{B'_1, B'_3\}$ is not connected. Thus, $\theta_{\alpha}^*(t_i) = \{B'_1 \cup B'_2 \cup B'_3\} = \{X\}$ and $\theta_{\alpha'}^*(t_i) = \{B'_1, B'_2, B'_3\}$.

Hence, even in the case when there is no chaining effect between adjacent blocks, $\theta_{\alpha}^*(t_i)$ need not refine $\theta_{\alpha'}^*(t_i)$ and $\theta_{\alpha'}^*(t_i)$ need not refine $\theta_{\alpha}^*(t_i)$.

In particular, $u_{\alpha}^* \not\leq u_{\alpha'}^*$ and $u_{\alpha'}^* \not\leq u_{\alpha}^*$.

Example 4.13. Let $\alpha > \alpha'$. Suppose $\theta_\alpha(t_{i-1}) = \theta_{\alpha'}(t_{i-1}) = \{B_1, B_2, B_3, B_4\}$, $d(B_1, B_2) = d(B_3, B_4) = t_i$, $d(B_1, B_3) = t_{i+1}$ and the rest of respective distances between these blocks are bigger than t_{i+1} . See Figure 4.

Since $\alpha > \alpha'$, we may assume, by condition ii, that there is an edge between B_1, B_2 and between B_3, B_4 in $G_\alpha^{t_i}$ but not in $G_{\alpha'}^{t_i}$. Thus, suppose $\theta_\alpha(t_i) = \{B_6, B_7\}$ while $\theta_{\alpha'}(t_i) = \{B_1, B_2, B_3, B_4\}$.

Now, we may assume that $\dim(F_{t_{i+1}}(B_1)), \dim(F_{t_{i+1}}(B_2)) < \dim(F_{t_{i+1}}(B_6))$ and $\dim(F_{t_{i+1}}(B_3)), \dim(F_{t_{i+1}}(B_4)) < \dim(F_{t_{i+1}}(B_7))$. Thus, we may also assume that, at t_{i+1} , for α' there is no edge between B_6, B_7 but for α there is an edge between B_1, B_3 . Therefore, $\theta_{\alpha'}(t_{i+1}) = \{B_6, B_7\}$ while $\theta_\alpha(t_{i+1}) = \{B_5, B_2, B_4\}$.

Hence, $\theta_\alpha(t_i)$ does not refine $\theta_{\alpha'}(t_i)$ and $\theta_{\alpha'}(t_i)$ does not refine $\theta_\alpha(t_i)$.

In particular, it is immediate to check that $u_\alpha \not\leq u_{\alpha'}$ and $u_{\alpha'} \not\leq u_\alpha$.

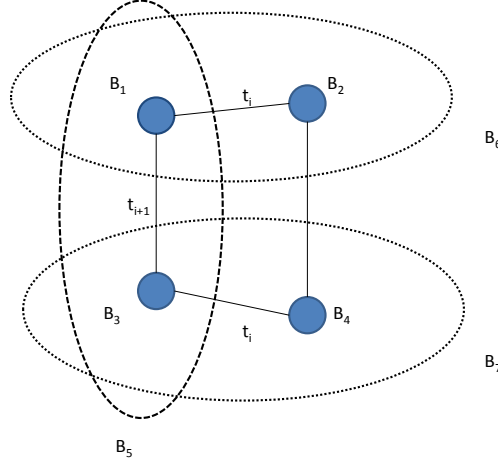


FIGURE 4. A bigger α does not imply a smaller ultrametric.

5. CONCLUSIONS

In the spirit of Kleinberg impossibility result we may consider A1 (the algorithm leaves ultrametric spaces invariant) and A3 (the distance between two points in the output is at least the minimal $\varepsilon > 0$ so that there exists a ε -chain between them in the input) as basic desirable conditions for any *HC* algorithm \mathfrak{T} . Thus, if we assume that \mathfrak{T} satisfies A1 and A3, then either \mathfrak{T} is exactly *SL* or else, condition A2 (adding points to the input will never make increase the distance between previous points in the output) is not satisfied. In particular, condition A2 is not satisfied by the algorithms defined to treat the chaining effects: $SL(\alpha)$ and $SL^*(\alpha)$.

Apart from this inevitable difference, we prove that the properties A1, A3, permutation invariance and richness are satisfied by $SL(\alpha)$ and $SL^*(\alpha)$ and also by the classical linkage-based algorithms, *SL*, *CL* and *AL*.

	SL	CL	AL	$SL(\alpha)$	$SL^*(\alpha)$
Permutation invariant	✓	✓	✓	✓	✓
Rich	✓	✓	✓	✓	✓
$A1$	✓	✓	✓	✓	✓
$A2$	✓	✗	✗	✗	✗
$A3$	✓	✓	✓	✓	✓
Semi-stable	✓	✓	✓	✓	✗
Stable	✓	✗	✗	✗	✗
Strongly chaining	✓	✗	✗	✗	✗
Completely chaining	✓	✗	✗	✗	✗
Weakly unchaining	✗	✗	✗	✓	✓
α -bridge-unchaining	✗	✗	✗	✗	✓

TABLE 1. Overview of the properties satisfied by the hierarchical clustering methods discussed in this work.

Their chaining and unchaining properties were studied in [9]. The stability properties of linkage-based methods were analyzed in [10]. The main results are summarized in Table 1.

REFERENCES

- [1] Ackerman M, Ben-David S, Loker D. *Towards Property-Based Algorithms Among Hierarchical Clustering Methods*. Neural Information Processing Systems Conference (NIPS 2010).
- [2] Ackerman M, Ben-David S, Loker D. *Characterization of Linkage-based Clustering*. (COLT, 2010).
- [3] Burago D, Burago Y, Ivanov S. *A course in metric geometry*. Graduate Studies in Mathematics. **33**, AMS, Providence, RI, (2001).
- [4] Carlsson G, Mémoli F. *Characterization, Stability and Convergence of Hierarchical Clustering Methods*. Journal of Machine Learning Research, **11** (2010) 1425–1470.
- [5] Gromov M. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, english edition, (2007).
- [6] Hughes B. *Trees and ultrametric spaces: a categorical equivalence*. Advances in Mathematics, **189**, (2004) 148–191.
- [7] Kleinberg J M. *An impossibility theorem for clustering*. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, NIPS. MIT Press (2002) 446–453.
- [8] Martínez-Pérez A, Morón M A. *Uniformly continuous maps between ends of \mathbb{R} -trees*. Math. Z., **263**, No. 3, (2009) 583–606.
- [9] Martínez-Pérez A. *A density-sensitive hierarchical clustering method*. arXiv:1210.6292v2 [cs.LG] (2013).
- [10] Martínez-Pérez A. *Gromov-Hausdorff stability of linkage-based hierarchical clustering methods*. arXiv:1311.5068 [cs.LG] (2013).
- [11] Zadeh R B, Ben-David S. *A uniqueness Theorem for Clustering*. Uncertainty in Artificial Intelligence. (UAI 2009).